COMPUTER-ASSISTED DRUG DESIGN

J. Phillip Bowen
University of Georgia, Athens, Georgia, U.S.A.

Michael Cory
GlaxoSmithKline, Research Triangle Park, North Carolina, U.S.A.

INTRODUCTION

Drug discovery and development are extremely time-consuming and costly processes. For every drug that reaches the market, there are more than 10,000 compounds synthesized, characterized, and tested for biological effects. Hundreds of millions of dollars are invested in the basic research and clinical studies which lead to its FDA approval and subsequent marketing. Traditionally, drugs have been "discovered" predominately through random or targeted screening efforts, followed by discrete structural changes in the molecule to optimize the properties responsible for the desired activity. With rapidly increasing costs, diminishing resources, more intense public scrutiny, greater government regulations, and higher expectations, this hit-and-miss approach is neither efficient nor economical.

The development of new medicines has been and continues to be a spectacularly successful scientific endeavor. While the traditional preparation and modification of complex pharmaceuticals has been a laudable achievement of synthetic medicinal chemists, the advent of computers has ushered in a new exciting approach to drug discovery. The new drug discovery activities, including that of computational chemists, have grown out of these technology changes. One of the clear trends of science has been the reduction of chemistry to particles in motion. Molecules undergo rapid translational, rotational, vibrational, and bending motions, which, in most cases, can be quantified using theoretical physics. Chemical reactivity may be attributed to these dynamic features of molecules. There is a direct relationship between molecular structure and biological activity. Through an intimate understanding of such molecular behavior, the medicinal chemist has the ability to gain profound insights into the most probable drug conformations, as well as the structural and energetic factors responsible for favorable drug–receptor interactions. With a combination of experimental data and computer-based information, the medicinal chemist is in a much better position to predict drug action. Unlike hand-held mechanical models, computer-generated structures add a new dimension to chemical perception by including the energy for a specific conformation defined by the Cartesian coordinates, which define the molecular array. Of course, other factors such as administration routes, absorption, distribution, metabolism, and elimination (the domain of pharmacokinetics) affect the activity of drugs and are increasingly being considered in the early phases of drug design.

The past twenty years have seen the development of a number of new approaches to drug design. The increased processing power of computers and the rapid display of computer-generated structures is simply amazing since the first writing of this article approximately a decade ago. These exciting developments over the last two or three decades have allowed the emergence of sophisticated computer programs specifically designed to assist medicinal chemists in developing new drugs. The combined application of molecular graphics, computational chemistry, as well as chemical and biological information is commonly called computer-assisted drug design (CADD). These computer-based approaches promise to fulfill a long-coveted goal of medicinal chemists: the prediction of biological activity prior to extensive laboratory synthesis and biological testing. One day it may not be unreasonable to expect medicinal chemists to design active molecular structures in a fashion analogous to the way engineers plan buildings, although the problems associated with biological problems are much more complex and frankly less understood. Computer-aided molecular modeling methods are still in their infancy, with exciting new methods and applications being reported at a staggering rate. Nevertheless, CADD approaches have already had a substantial impact on pharmaceutical design, discovery, and development, moving from the realm of after-the-fact rationalizations to a more beneficial predictive role.

This article will provide a brief introduction to computational chemistry, molecular modeling, and CADD theory. Clearly, it is impossible to cover in this short review all of the useful and/or interesting CADD approaches. Although this is relatively a new field of

research, there are many examples from which to choose. Today, there are a number of pharmaceutical agents in various stages of clinical trials and/or on the market in which computer-based methods were used. Research examples have been included to illustrate more completely the conceptual points. Some of the fundamental methodologies are discussed, including the important underlying mathematical formalisms. In addition, historical perspectives have been incorporated. Finally, the ever-changing hardware and software trends are discussed.

METHODS

Molecular Modeling

Molecular modeling of organic compounds has a long history. John Dalton (1) had a set of wooden spheres, representing atoms, drilled to receive rods representing bonds. These original ball-and-stick models are on display at the London Museum of Science and date to 1810. The most widely known recent application of molecular modeling to the solution of a significant organic chemistry structural problem is the Nobel prize-winning work of J. D. Watson and F. H. C. Crick for the development of their model for the structure of DNA. Initially, Watson and Crick used homemade molecular models cut by hand from cardboard. Later, as they improved their understanding of the structures and the conformations of the bases in the nucleic acids, they had more exact models machined from metal. At the time of their work (mid 1950s) controversy existed over the correct tautomeric forms for the heterocyclic bases. Based on old experimental results, the first Watson and Crick models had the incorrect tautomers. After building models with the correct tautomeric forms of the bases, they quickly arrived at a conformation for DNA that not only matched experimental data but also suggested a method by which DNA could be responsible for cellular replication (2, 3).

In the two decades following this landmark work, various types of hand-held models became commercially available, and they have been used extensively by medicinal chemists. The wooden ball-and-stick models frequently used by high school students, as well as the metal Dreiding and Kendrew (4) models (not to mention the plastic varieties), are still quite popular in organic chemistry laboratories. Before the availability of high-resolution computer graphics systems, X-ray crystallographers depended heavily on metal models, including the Kendrew type, for converting computer-plotted electron density maps into physical models that could be used to study proteins.

Linus Pauling, in his early studies of the conformations of the *alpha* helix in proteins, suggested it would be helpful to obtain solid molecular models that would allow modeling of large pieces of a polypeptide structure. This suggestion eventually led to the CPK (Corey, Pauling, Koltun) (5) spacefilling models which are made from plastic and are designed to effectively represent the true size of atoms. These models were designed under the guidance of an IUPAC committee. Their quality and popularity led to their commercial availability (6). Many laboratories used large quantities of these models to assist in drug design and the understanding of molecular shapes. A few laboratories attempted to use such models for the study of proteins and nucleic acids. Due to their tactile properties, these models are excellent teaching tools and led to an easy understanding of problems in stereochemistry and molecular shape. The mechanical stress on the plastic parts, however, resulted in compromising the integrity of models of large molecules. It turned out that the models for large molecules were less useful than computer representations. Further complications arise when attempts were made to simulate ligand–protein (or drug–receptor) interactions. Importantly, there is no possibility of determining accurately the energy differences among various conformations.

The display of molecular representations on a computer screen has been an objective of many computer scientists over the years. Significant decreases in the cost of computing coupled with dramatic increases in the speed and quality of computer displays have led to the increasing use of computer graphics displays for molecular modeling. Graphics displays have grown from the early ball-and-stick ORTEP (7) plots of small X-ray crystal structures to interactive displays that show spacefilling models of entire proteins. Computer-generated molecular displays are not only integral parts of structural biology and structural chemistry research, but they have also moved into the realm of advertising, as seen on the covers of scientific journals and in the many pharmaceutical media advertisements. In other words, computer-generated images not only impart specific and sometimes subtle chemical and biological information not possible to obtain with hand-held models, but they also tend to represent state-of-the-art methods. Most of these developments were driven by the rapidly growing computer-assisted design market, rather than the much smaller chemistry market.

The growth of visualization applications has been boosted by improved computer graphics displays and increasing computational chemistry capabilities, as indicated above. The initial growth of the market was induced through the rapid acceptance of computer-assisted

design in the aircraft manufacturing industry. Until about 1983, sophisticated molecular modeling software was the domain of some academic scientists; most of the advances took place in X-ray crystallography laboratories. The most sophisticated of these displays allowed full three-dimensional transformation of the coordinate display by turning a dial. Custom-designed computer molecular modeling and display software, now available from multiple vendors, allow the interactive manipulation of the display through a joystick, mouse, dial, and/or keyboard commands.

The use of computer-generated molecular models has many advantages over the more traditional manipulation of physical molecular models. Since the molecules are represented in the computer as sets of three-dimensional (3-D) atomic coordinates, a large number of molecules can be handled simultaneously. A feat which is essentially impossible using mechanical models. Any number of computer-generated images can be superimposed and analyzed. These techniques have provided medicinal chemists with information about the relative size and shape of sets of molecules with similar biological properties. The superimposition of different conformations of the same molecule can provide information about the relative space a molecule can occupy. Since the computer maintains a set of coordinates for each of the atoms in each molecule, information about bond lengths, bond angles, and torsion angles of a model can be rapidly obtained. Additionally, nonbonded and direct through-space distances, so important in defining the shape of receptor sites, can also be readily obtained and compared. Moreover, the information can be stored, retrieved, and re-analyzed at later times.

From the preceding discussion, one extremely important aspect of molecular modeling which cannot be overstated: visualization of molecules. When molecular interactions between a drug and its receptor are displayed, manipulated, and rotated on a graphics terminal, researchers can identify important and unimportant molecular interactions. It is through such computer exercises that medicinal chemists can confirm, reject, or develop hypotheses regarding the molecular origins of biological activity. Perhaps one major drawback to computer-generated structures is the fact that one is unable to handle them physically, which could be an artifact of previous experiences with hand-held molecules in sophomore organic chemistry classes. Interestingly, there is a growing trend in organic chemistry education to move toward computer graphic representations of organic structures. On computer graphics terminals, to give the illusion of three-dimensional visualization, a technique called *depth cueing* is used to shade structures in such a way that objects further away from the viewer are slightly dimmer, while objects that are closer appear brighter. This optical effect works reasonably well. Another stratagem that creates the illusion of three-dimensional perception is to have the molecular structure slowly rotate about its center of mass or slowly rock back and forth. Current technology allows breathtaking stereo viewing of molecular display on desktop computers. An example of this is the crystal-eyes system (http://www.stereographics.com/).

There continues to be important developments in the molecular modeling area. The first is the development and use of algorithms that describe the surface area of molecules. Software utilizing these algorithms allows the comparison of computed surface areas. One of the first algorithms to be widely implemented in the study of molecule surfaces is that of Connolly (8). In it, a sphere of a given radius (usually the radius of a water molecule) is "rolled" over a defined surface, usually the van der Wäals surface. Wherever the sphere touches the macromolecule, a new contact surface is generated. The display of this surface allows the medicinal chemist to see what should be considered the *solvent accessible* surface. An alternative approach, much more efficient in computational and display time, is to place a dot surface over the molecule and to compute the surface using expanded atomic radii. The more recent implementations of this algorithm display the electrostatic density at the surface. This provides the chemist with a picture of what charge might be exposed to solvent. As applied to drug design, the *solvent excluded* surface allows an indication of the topology of a receptor site. Plotting physicochemical functions, such as the electrostatic potential, on the surface of this model indicates the forces acting upon a molecule that binds to that surface. Today, it is possible to display transparent molecular surfaces rapidly, which gives the impression that the molecular skeleton is suspended in a gelatin-like volume.

The second major change, interrelated to the discussion above, is related to the decreased cost and increased speed of computation. Raster computer displays generated widespread interest in the development of display algorithms for calculating high resolution spacefilling representations of molecules. Many algorithms, motivated in part by the advances in computer technology, were developed to give rapid, high resolution CPK-like displays of molecular surfaces. Most of these systems provide a static display computed from a specific molecular orientation. Real-time, three-dimensional manipulation of these "CPK" models was demonstrated successfully on some systems (9) and is now standard.

Computational Chemistry

Overview

Much of the recent progress in drug design has been based on the fundamental understanding of drug-receptor interactions and the increased availability of high quality structural data. Quantitative correlations have been made between biological activity and molecular properties. Computer experiments offer complimentary means to acquire essential information that may be quite difficult or impossible to obtain empirically. Of course, these theoretical approaches need to be properly calibrated to existing experimental data to insure that if there are any systematic differences they are clearly understood at the outset. No one is seriously suggesting that computer analysis will replace experimental work. The mathematical models presently used are just that—mathematical models approximating complex physical and biological properties. Traditionally, a great deal of attention has been focused on the structural fit between receptors and ligands, like keys within locks. Today, this idea of a lock-and-key analogy is a well-recognized exaggeration since small molecules and receptors are not static structures but constantly changing to adopt new conformations in response to their environment. In considering pharmacodynamic interactions, favorable electrostatic interactions are critically important. Charge density and the complimentarity of charges between ligand and receptor are readily accessible to calculation, althogh there are several different ways to generate charges. Important quantitative structure–activity relationships (QSAR) between nontraditional parameters such as HOMO–LUMO interactions and bond orders have been correlated with biological activity.

Researchers involved in drug design have been criticized for focusing too extensively on the interaction energies governing ligand–receptor binding while neglecting important solvation and desolvation effects and pharmacokinetic effects. These other aspects of drug availability, as well as toxicity, are more complex to model and consequently more difficult to predict. Progress, however, is being made. Until much more research verifies the accuracy of pharmacokinetic and toxicity models, energy-based approaches focused on drug–receptor fitting will dominate. Alternative chemical information based methods will play increasingly important roles in drug design.

Quantum mechanics

History and concepts: In the late nineteenth and early twentieth centuries, the applications of classical mechanics failed to describe accurately blackbody radiation or atomic spectra (10a, 10b). For nearly two centuries, Newtonian mechanics had been the cornerstone of physics, essentially unassailed and unfailing. According to the prevailing attitude among most physicists of the late nineteenth century, all physical theories had been discovered (10a, 10b). The only remaining challenge for future generations of scientists was to refine minor details and extend the accuracy to an additional decimal place. Wein in 1896 and Rayleigh in 1900 had tried to explain blackbody radiation using standard classical methods. Every attempt, however, by theoreticians of the day failed to reproduce all of the available empirical data. This failure of classical physics has been termed the "ultraviolet catastrophe." It was not until Planck made his revolutionary proposition, that the oscillators themselves needed to be quantized, that a consistent mathematical solution emerged.

Based on the Einstein wave-particle solution for the photoelectric effect, De Broglie was the first to suggest that if particle characteristics were associated with electromagnetic radiation, then matter should also have wavelike characteristics. De Broglie derived a simple relationship (Eq. 1) correlating the wavelike characteristics, λ, with momentum, p. Large objects would have a vanishingly small wavelength. The wavelike characteristics of matter would only manifest themselves significantly at the atomic and subatomic levels since the wavelength is inversely proportional to the mass. The de Broglie equation was used to predict exact diffraction patterns which were later verified experimentally by Davisson and Germer.

$$\lambda = \frac{h}{p} \tag{1}$$

Since matter had wavelike properties, it could argue that there must be an underlying fundamental wave equation governing the behavior of matter. Erwin Schrödinger (11a, 11b, 11c) formulated the celebrated differential equation, which, for chemists, is the basic molecular expression of quantum mechanics. The idea being that if one could solve the Schrödinger equation, then all of the physical properties of a molecular system (or any system of any size for that matter) would be known. Chemistry, and ultimately biology, could be reduced to solving mathematical equations. Another more general formulation of quantum theory, using matrix algebra, was developed independently by Heisenberg and Dirac (12). Both methods have been demonstrated to be equivalent mathematical representations (13a, 13b).

The concept of spin falls naturally out of the matrix mechanics approach, but it is an ad hoc addition to wave mechanics.

To date, quantum theory, despite its many peculiar "nonclassical" microscopic aberrations, has been used effectively to explain experimental observations and to predict accurately physical effects in advance of the experiment. Interestingly, many of the founders of quantum mechanics later rejected it, primarily because it was a nondeterministic theory. The break from their classical mechanics view of the universe appears to have been too severe for them to accept. But a new generation of physicists was prepared to embrace the quantum mechanics and apply the methods to chemical structures. With a series of novel concepts and observations, physics and chemistry were changed forever. From the brilliant work of Schrödinger, de Broglie, Heisenberg, Dirac, Mulliken, Pauling, and many other pioneers, the new quantum chemistry was born.

Methodology: Unquestionably, the application of quantum mechanics to chemical bonding has revolutionized scientific thinking. In fact, the modern theoretical framework of chemistry rests on quantum physics. In principle, the Schrödinger equation may be solved for any chemical system. No prior knowledge of any analogous or related system is necessary. Exactly solvable problems are rare, due to the mathematical complexities; recourse must then be made to approximate methods, and many powerful approaches have been devised. Generally, approximate solutions must suffice for the size of molecules of pharmaceutical interest.

The Schrödinger equation is the starting point for molecular problems (13a, 13b). The symbol H is a differential operator called the Hamiltonian operator, which is analogous to the classical Hamiltonian, inasmuch as it is a sum of kinetic and potential energy terms. E is the total energy for the system. The wavefunction Ψ depends on the position of all the particles comprising the system. Born (10a, 10b) proposed that $|\Psi|^2$, and not Ψ, is a measure of the probability distribution of the particles within a molecule. The Schrödinger equation is an eigenfunction, which in its most general form is written as follows (Eq. 2):

$$H\psi = E\psi \tag{2}$$

All quantum chemical calculations rely on a very important assumption called the Born–Oppenheimer approximation, which treats nuclear and electronic motions separately (Eq. 3), and it is discussed in most elementary quantum mechanics textbooks.

$$(H_{\mathrm{el}} + V_{NN})\psi_{\mathrm{el}}(r, R) = E\psi_{\mathrm{el}}(r, R) \tag{3}$$

The nuclei, being much heavier than electrons, are considered fixed in space relative to the faster moving electrons. V_{NN} is a function describing the nuclear potential energy and may be factored out of the electronic expression. This uncoupling of electronic and nuclear motion is critical to quantum mechanical formulations. Nevertheless, the Born–Oppenheimer approximation may not be accurate under certain circumstances. Once a nuclear configuration has been determined, the Schrödinger equation is then solved to give the electronic energy for a particular nuclear configuration. The process may be repeated over every conceivable nuclear arrangement. A potential energy surface can then be mapped. The most stable nuclear–electronic configuration has the lowest energy, and it is referred to as the *ground state*. Promotion of electrons to higher energy states, referred to as *excited states*, usually is not of importance in most drug design applications, although there are exceptions. The vast majority of energy computations related to pharmaceutical agents and their design have been carried out on ground state electronic configurations.

In general, quantum chemical calculations may be divided into three broadly defined subdisciplines based on the approaches taken to solve the Schrödinger equation: 1) semiempirical-based methods, 1, 2f) ab initio-based methods, and (c) density functional theory. John Pople, ab initio methods, and Walter Kohn, density functional theory (DFT), received the 1998 Nobel Prize in chemistry for their pioneering work in computational quantum chemistry.

a. *Semiempirical calculations:* Exact solutions to the Schöringer equation are limited to special cases: a particle in a box, a particle on a ring, a harmonic oscillator, the hydrogen atom, and the hydrogen molecule ion, etc. In an effort to apply quantum mechanics effectively to chemical systems of more practical importance, several simplifying approximations have been made and refined over the years. Early work by Pople and co-workers (14a, 14b, 14c, 14d, 14e, 14f, 14g) led to an approximation called Complete Neglect of Differential Overlap (CNDO). The central idea behind this approximation was to eliminate certain sets or families of two-electron integrals of the following expression (Eq. 4). The solution of these expressions are the most complicated mathematical steps in quantum chemical calculations,

especially when integration is over four different atomic centers a, b, c, and d. The omitted integrals could then be accounted for by adding empirical equations and parameters. Years later, Dewar developed a series of methodologies that approach chemical accuracy (15). These methodologies have been incorporated into a series of computer programs, MINDO (modified intermediate neglect of differential overlap; 16), MINDO/3 (17a, 17b, 17c, 17d, 17e), MNDO (moderate neglect of differential overlap; 18a, 18b, 18c, 18d, 18e, 18f), and AM1 (Austin Method 1) (19). Correlations between drug activity and molecular orbital energies and coefficients have been made (20a, 20b, 20c). These and similar approximations are classified as semiempirical calculations.

$$\Psi_a^*(1)\Psi_b(1)\left(\frac{1}{r_{12}}\right)\Psi_c^*(2)\Psi_d(2) \tag{4}$$

b. *Ab initio calculations:* The second quantum mechanical approach is based on what is referred to as the ab initio method (21). As the name implies, ab initio calculations use a nonparameterized method. Strictly speaking, these calculations also have a number of simplifying approximations; but they are much more demanding computationally, as no classes of integrals are eliminated, and consequently are much more expensive. There have been numerous reports and reviews in the literature on ab initio and semiempirical calculations, as well as detailed comparisons of the two methodologies (21a, b).

Usually the Ψ function is formulated from Molecular Orbital (MO) theory, where one-electron orbitals are used to approximate the true many-electron wavefunction. It is assumed that each molecular orbital is a truncated linear combination of n one-electron functions, known as the basis set, and are primarily atomic orbitals. This approach is known as the linear combination of atomic orbital (LCAO) method (23a, 23b). The actual functional form of the wavefunction Ψ determines the complexity of the calculation. Boys (24) originally suggested that Gaussian Type (GT) functions may be used to describe Slater Type (ST) functions, which in turn are approximations of the true hydrogenic orbitals. The use of GT functions greatly reduces computational time because they can be solved explicitly. Since GT functions, unlike ST orbitals, do not have a cusp at the origin, they are less accurate. In addition, as distance from the origin increases the GT functions decay too quickly. Nevertheless, despite these limitations, a linear combination of GT functions can be made to fit an ST function.

Pople found that a minimal basis set was composed of a linear combination of three GT functions; they are referred to as STO-3G (25a, 25b, 25c, 25d). Pople also found that more accurate representations may be achieved with more sophisticated GT combinations. The next level of theory splits the outer shell basis functions into two parts, appropriately known as split-level basis sets. In the early 1990s, 3–21G calculations became the new standard (26a, 26b, 26c), replacing the STO-3G formulation. Presently, more complex basis sets have become the new standard. Additional accuracy was achieved by adding polarization functions, which may be denoted by (*). Other conventions are also used: for example, 6–31G* calculations, the new standard, treat the core electrons with six primitive basis functions (27a, 27b, 27c). The outer shell electrons are divided into two parts, which are composed of a set of three GT functions and one GT function, respectively. Finally, a set of six d polarization functions has been added for non-first-row elements (21, 28). The split basis sets and polarization functions add flexibility to the molecule such that the atomic orbitals can more accurately respond to their environment. Polarization of hydrogen atoms is achieved by adding a set of three p orbitals, denoted by a second * (e.g., 6–31G**). The next level of accuracy divides the outer electrons into three sets of Gaussians (29). With the advances in computers, 6–311G* calculations are more economical and becoming more popular.

Quantum mechanical calculations are carried out using the Variational theorem and the Hartree–Fock–Roothaan equations (13a, 13b, 21, 30a, 30b, 30c). Solution of the Hartree–Fock–Roothaan equations must be carried out in an iterative fashion. This procedure has been called self-consistent field (SCF) theory, because each electron is calculated as interacting with a general field of all the other electrons. This process underestimates the electron correlation. In nature, electronic motion is correlated such that electrons avoid one another. There are perturbation procedures whereby one may carry out post-Hartree-Fock calculations to take electron correlation effects into account (31a, 31b, 31c, 31d, 31e). It is generally agreed that electron correlation gives more accurate results, particularly in terms of energy.

c. *Density functional theory:* Density functional theory (DFT) is the third alternative quantum mechanics method for obtaining chemical structures and their associated energies (32). Unlike the other two

approaches, however, DFT avoids working with the many-electron wavefunction. DFT focuses on the direct use of electron densities $P(r)$, which are included in the fundamental mathematical formulations, the Kohn–Sham equations, which define the basis for this method (33). Unlike Hartree–Fock methods of ab initio theory, DFT explicitly takes electron correlation into account. This means that DFT should give results comparable to the standard ab initio correlation models, such as second order Møller–Plesset (MP2) theory.

In some ways, DFT may be more easily understood. According to the Kohn–Hohenberg theorem (34), the energy is minimized when the calculated and true electron densities are equal. One important consequence of DFT calculations is that their accuracy corresponds to standard post Hartree–Fock methods. Another consequence of DFT calculations of more practical importance is the reduction in the computation time required to complete a calculation. The time required to complete Hartree–Fock calculations is a function of the number of electrons in the system being examined, and it is proportional to n^4, where n represents the number of electrons. DFT calculations, also a function of the number of electrons in the system, are proportional to n^3. (Contrast the time for quantum chemical calculations with those of molecular mechanics calculations, discussed later. In molecular mechanics, the time scales to m^2, where m is the number of atoms, not the number of electrons. Molecular mechanics is an attractive and accurate alternative for larger molecular systems where it is impossible to use quantum mechanics based methods.) For systems of interest to medicinal chemists, the differences in computer time between DFT or ab initio calculations may be the most important factor in selecting which method is used. Although there are more examples of ab initio calculations in the literature and consequently more experience with their accuracy and limitations, there is a growing number of DFT calculations being reported.

Molecular mechanics

History and concepts: A complementary approach for molecular structure calculations is available, and it is referred to as the molecular methanics or force field method; it is also known as the Westheimer method (35). In 1946, twenty years after the impressive development of quantum theory, three papers appeared in the literature which applied *classifical* mechanical concepts to problems of chemical interest (36–38). Westheimer investigated the racemization of some optically active biphenyl derivatives. His work demonstrated the potential usefulness of molecular mechanics. The other two papers were attempts to tackle more complex problems.

The fundamental concepts of molecular mechanics are deeply rooted in traditional chemical thought, which takes the view that molecules may be considered as a series of masses held together by elastic or harmonic forces (somewhat like balls fastened together by weightless springs). The idea central to molecular mechanics theory, concisely stated, is that any deformation of the ball-and-spring model from its *natural* bond lengths (1_0) and angles (θ_0) results in a strain. These deviations from geometric ideality are reflected in a correspoding increase in energy for the model. A set of classical equations are used to describe the motion and corresponding energy of the spring. Hence, with the wealth of chemical information found in the literature, parametrization of these *classical* equations results in the accurate reflection of moelcular behavior—that is to say, the potential used to describe the energy surface are generally referred to as the force field. In physics the term *field* traditionally is given to the continuous distribution of some "condition" prevailing through space.

An isolated molecule may presumably adopt various conformations in response to specific intramolecular forces. In general, the total energy of a molecule (E_{total}) may be regarded as the summation of stretching (E_s), bending (E_b), torsional (E_{tor}), van der Waals (E_{vdw}), and electrostatic (E_{ele}) energy components. For more advanced force fields it is crucial for structural accuracy to include various cross-terms ($E_{cross\ terms}$), e.g., stretch–bend, torsion–stretch, and torsion–bend, etc., are added to reflect the fact that motions and interactions are coupled to one another. Additional terms may be added, such as hydrogen bonding potentials. One of the fundamental theorems of molecular mechanics is the division of the total energy into separate, readily identifiable parts, as indicated by Equation (5). As discussed previously, deviations from the *natural* geometry is accompanied by an increase in the total energy (E_{total}) of the molecule. The total energy may be viewed as a summation of various energy terms that medicinal scientists feel comfortable in using an visualizing (Eq. 5). A second tenet of molecular mechanics theory concerns the transferability of the parameters used in the individual potential energy terms of Eq. 4. In other words, the parameters used are transferable from one molecule to the next. The continued success of molecular mechanics in treating a wide variety of compounds has justified these assumptions. When several molecular structures are to

be considered together, however, one must obviously think about intermolecular forces as well.

$$E_{total} = E_s + E_b + E_{tor} + E_{vdw} + E_{elec} + E_{cross\ terms} + \cdots \quad (5)$$

Some theoretical purists tend to view molecular mechanics calculations as merely a collection of empirical equations or as an interpolative recipe that has very little theoretical justification (39). It should be understood, however, that molecular mechanics is not an ad hoc approach (35). As previously described, the Born–Oppenheimer approximation allows the division of the Schrödinger equation into electronic and nuclear parts, which allows one to study the motions of electrons and nuclei independently. From the molecular mechanics perspective, the positions of the nuclei are solved explicitly via Eq. 2. Whereas in quantum mechanics one solves, Ψ, which describes the electronic behavior, in molecular mechanics one explicitly focuses on the various atomic interactions. The electronic system is implicitly taken into account through judicious parametrization of the carefully selected potential energy functions.

Methdology: The fundamental molecular mechanics formulation may be derived from the expansion of the potential energy function $E(q_1,q_2,q_3, \mathrm{H}, q_1)$ in a Taylor series about the equilibrium position q_0, where q represents generalized coordinates (40), yielding Eq. 6:

$$E = E_0 + \sum_{i=1}^{n}\left(\frac{\partial E}{\partial q_i}\right)_0 \Delta q_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{\partial^2 E}{\partial q_i \partial q_j}\right)_0 \Delta q_i \Delta q_j + \frac{1}{6}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{o}\left(\frac{\partial^3 E}{\partial q_i \partial q_j \partial q_k}\right)_0 \Delta q_i \Delta q_j \Delta q_k + \ldots \quad (6)$$

The first term in the expansion is a constant. Since there is the freedom to select a point of zero potential energy E_0 can be defined as zero. The second term is the negative of the restoring force, which is equal to zero at the equilibrium position (Eq. 7):

$$-F_i = \left(\frac{\partial E}{\partial q_i}\right)_0 = 0 \quad (7)$$

Hence, the first nonvanishing term in the Taylor series expansion is the third term, which is a quadratic expression. This third nonvanishing term corresponds to a Hooke's law potential in the limit of small vibrations. The series may be truncated after the third term to provide the following potential energy expressions (Eqs. 8–10):

$$E_s = \frac{1}{2}\sum_{i=1}^{n} k_q (q_i - q_0)^2 \quad (8)$$

$$E_\theta = \frac{1}{2}\sum_{i=1}^{n} k_\theta (\theta_i - \theta_0)^2 \quad (9)$$

Additionally, cross terms such as the stretch–bend potential, shown below in Eq. 10, may be included in the force field equation to give better agreement between experiment and calculation.

$$E_{\text{stretch–bend}} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} k_\theta (q_i - q_0)(\theta_j - \theta_0) \quad (10)$$

The stretching, bending, and various cross-term potentials [specifically illustrated with a stretch–bend effect in Equation (10) above] naturally arise from the Taylor series expansion. Cross-terms have been successfully added to various force fields to give more structural flexibility. For example, when intramolecular angles are compressed, experimental studies indicate that the bonds stretch to relieve the strain resulting from the angle bending. Mathematically, in a valence bond force field, this may be incorporated into a force field potential through cross terms. In the case of angle compression and bond stretching, a stretch–bend term may be introduced (see Eq. 10). There are other ways to treat these effects. Explicit terms between 1,3 nonbonded atoms may also be incorporated, and are called Urey–Bradley force fields. The MM2, MM3, and MM4 force fields (41a, 41b, 41c, 41d, 41e, 41f, 41g, 41h, 41i, 41j, 41k) uses stretch–bend cross terms to mimic this type of molecular behavior. In the latest formulation of molecular mechanics treatments, Allinger and co-workers have incorporated a torsion stretch term to represent quantitatively the bond stretching that occurs in strained molecules. It has been shown that some cross terms are more important than others, which means that some can be neglected without drastically affecting the calculated structures. Additionally, certain cross terms are known to affect calculated vibrational spectra.

Higher terms from Eq. 6 may also be added to the simple quadratics to furnish better representations of chemical bonding. Nevertheless, these energy terms alone do not adequately describe molecular behavior.

The third energy term of Eq. 5 arises from pairwise nonbonded interactions summed over every pair i and j. Two opposing forces play a role in this energy term. For relatively long distances, two atoms attract each other due to London dispersion interactions, which are proportional to r^{-6}. As the two atoms come into close proximity, they

exert a mutual repulsion known as van der Waals repulsion. the Leonard-Jones 6–12 potential function [Eq. 11] describes this behavior (42), originally derived for the noble gases, and is most often used in simple force fields. Quantum mechanical calculations indicate that the repulsion part of the curve may in fact be somewhat overestimated. Some force fields use a 6–10 potential to reduce the steepness or hardness of the curve as two atoms approach one another, which is perhaps most commonly used for cases of hydrogen bonding. The use of an exponential term instead of the r^{-12} term in force field equations better reproduces experimental data for organic structures and is more consistent with quantum chemical calculations. Potential exponential functions have been known for some time. In fact, MM2, MM3, and MM4 use a modified Hill equation to determine the nonbonded interactions.

$$E_{vdw} = \sum_{i|j} \sum \left[\frac{A}{r^{12}} - \frac{B}{r^{6}} \right] \tag{11}$$

It has long been known that molecules do not rotate freely. In 1891, Bischoff proposed that ethane preferred a staggered conformation and that restricted rotation occurred in substituted ethanes (43). Christie and Kenner first demonstrated restricted rotation in 1922 by resolving 2,2′-dinitrophenyl-6,6′-dicarboxylic acid into optically active isomers (44). Pitzer showed that the calculated and observed entropies for ethane were identical if restricted rotation was considered (45a, 45b). The phenomenon of restricted rotation appears to be ubiquitous and has stimulated intense interest and research.

There has been much speculation concerning the origin of the internal energy barrier (46). Three explanations which have been proposed: 1) repulsion between the bonds not covered by the van der Waals repulsion; 2) correction for the anisotropy of van der Waals repulsion; and 3) various quantum mechanical suggestions.

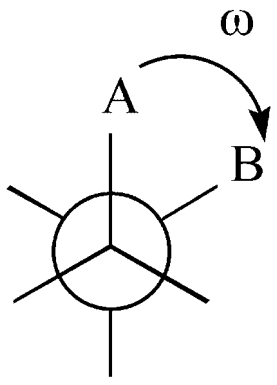


Fig. 1 Dihedral angle representation.

The barrier to internal rotation must be included in the force field. Usually, the internal rotational energy is a function of the torsion angle ω. (See Fig. 1.) Klyne and Prelog have defined the torsional angle ω for the connected atoms A—C—C—D, as depicted previously. Rotation of A towards D along the shortest arc, viewing the molecule along the C—C axis, is defined as a positive value when clockwise.

The torsional potential may be expressed as a truncated Fourier series, as shown in Eq. 12, where the summations is over all n unique bond dihedrals (torsions), ω. Usually, the first three terms are sufficient to describe the torsional energy adequately. Higher terms may improve the fit between experimental data and calculation. Inclusion of V_1 and V_2 terms has improved the agreement between experiement and calculations (35a, 47b, 47c). Radom, Hehre, and Pople (48) had given a physical explanation to the torsional terms. The onefold term, V_1, is a dipole–dipole interaction. The twofold term, V_2, arises from hyperconjugation. The threefold term, V_3, has a steric origin.

$$E_{\text{tor}} = \sum_{i=1}^{n} \left[\frac{1}{2} V_{i,1}(1 + \cos\omega) + \frac{1}{2} V_{i,2}(1 - \cos 2\omega) + \frac{1}{2} V_{i,3}(1 + \cos 3\omega) + \ldots \right]$$

Other related Fourier expressions may include phase angles, which may also be used to describe potential energy curves; this is the approach employed in AMBER (49a, 49b, 49c, 49d). See Eq. 13, where ϕ is the dihedral angle, V^{Φ} is the force constant, n is the multiplicity, and δ is the phase angle. These Fourier series expressions allow the incorporation of additional experimental or quantum mechanical phenomena into force field calculations.

$$E_{\text{tor}} = \sum_{i=1}^{n} V_{\phi} \left[1 + \cos(n\phi - \delta) \right] \tag{13}$$

Electrostatic interactions play a crucial role in the binding associated with drug–receptor interactions. The electrostatic part of any molecular mechanics potential energy expression is extremely important for many types of calculations beyond drug–receptor considerations. Interestingly, consensus opinion on the best formulation has not been reached. It appears that two alternative, but ultimately equivalent, representations may be used: 1) point charge–point charge and 2) dipole–dipole

interaction models. Most force fields utilize the former, explicitly assigning point charges to each atom in a molecule and relying on Coulombic interaction terms, where the energy varies between two point charges, q_1 and q_2, by r^{-1}. Mathematically, this formulation requires the evaluation of square root terms. Originally when applied to large biopolymers, the electrostatic calculations were be quite time consuming. An approximation which successfully circumvents this problem sets the dielectric constant equal to r (50a, 50b): With this stratagem, q_1 and q_2 very as r^{-2}, and this in effect dampens the charge interactions. The Allinger force fields and the many variants treat electrostatic interactions with a dipole–dipole interaction equation, where the dipole energies vary by r^{-3}. For charged structures, there must be dipole-charge and charge–charge terms. The charges or dipoles are not static, and the recipes are influenced by neighboring groups. Ideally, terms or methods which consider the polarizability of groups are increasingly being used (51). Progress over the years continues.

The treatment of charges in molecular mechanics calculations remains a central question. The exact computational scheme used to obtain charges has been under considerable review. One obvious way is to obtain the charges from quantum mechanics calculations. Since quantum mechanically derived point charges are based on various population analyses, the level of accuracy required and how the charges should be distributed between the atoms have not been universally accepted. One final check on the charge distribution in a molecule is the comparison of the calculated dipole moment with experimentally determined dipole moments. If there is a reasonable charge distribution, based on calculated charges, then the calculated and experimental dipole moments should be in close agreement.

As indicated earlier, molecular mechanics has also been described as the empirical method. Most of the parameters and equations are designed to reproduce experimental data (for example, molecular geometry, dipole moments, and heats of formation). Perhaps the term *empirical method* is becoming outdated. In many cases, high level ab initio (or DFT) calculations are being used to augment sketchy or nonexistent experimental data such as charge distribution, conformational energies, and rotational barriers (52a, 52b, 52c, 52d, 52e, 52f). Recent work by several groups has used a combination of experimental and theoretical work to parameterize transition state force fields (53a, 53b, 53c, 53d, 53e).

Geometry optimization: Once the three-dimensional Cartesian coordinates and necessary parameters (such as atom and/or bond types) have been established for a molecular structure, the next step is to determine the energy-optimized structure. Various molecular mechanics programs have geometry optimization algorithms built into them, and they are capable of locating an atomic arrangement at an energy minimum. The initial structure has an energy associated with it based on the particular force field. Every atom has a net force acting upon it. The objective is to relieve as many of the forces as possible, such that the derivative of the potential energy equation is zero.

A variety of methods have been developed for these purposes. Most of the modern methods involve differentiation of the analytical functions, whereas earlier methods used numerical techniques to reach the lowest point (bottom) of the potential energy well. These methods assume the potential energy function is well behaved and differentiable. On the one hand, the simplest approach is called *steepest descent* (54a, 54b, 54c). In this method, displacement steps k_i are taken in the direction parallel to the net force acting on the atom. Since the displacement vector is parallel to the force, which in turn is equal to the negative of the gradient of the potential, the direction is straight downhill. Steepest descent is relatively efficient when the geometry is poor, i.e., far from the minimum. As the minimum is approached, however, the derivatives become smaller and smaller. Consequently this algorithm has reduced efficiency in these cases.

On the other hand, conjugate gradient methods (55a, 55b) are more effective in locating the minimum energy structure. In this approach previous optimization information is utilized. The second and all subsequent descent directions are linear combinations of the previous direction and the current negative gradient of the potential function. Conjugate gradient methods are generally more efficient than steepest descent. In general, for an N-dimensional quadratic surface, the minimum will be located in N steps.

The Newton–Raphson approach is another minimization method (56a, 56b). It is assumed that the energy surface near the minimum can be described by a quadratic function. In the Newton–Raphson procedure the second derivative or **F** matrix needs to be inverted and is then used to determine the new atomic coordinates. **F** matrix inversion makes the Newton–Raphson method computationally demanding. Simplifying approximations for the **F** matrix inversion have been helpful. In the MM2 program, a modified block diagonal Newton–Raphson procedure is incorporated, whereas a full Newton–Raphson method is available in MM3 and MM4. The use of the full Newton–Raphson method is necessary for the calculation of vibrational spectra. Many commercially available packages offer a variety of methods for geometry optimization.

In many cases, low-energy conformations are sought, although the drugs bound to receptors are required (and usually are not) in their lowest energy state for the isolated molecule. All of the algorithms, described above, have been almost exclusively designed to find minima, but not necessarily the global minimum. When an energy optimization has been carried out for a complex molecule, it is impossible to tell whether or not the absolute lowest energy conformation has been calculated. In general, programs are not designed to go over energy barriers. The energy-minimized structure depends on the starting coordinates initially used. So if the coordinates corresponding to the boat conformation of cyclohexane were used as the starting point, then the force field optimizer would seek out the closest energy well. In this case, the final geometry would correspond to a twist-boat structure, a minimum, but not the chair form, the global minimum.

As indicated, there is no assurance that the *bioactive* conformation is necessarily the *global* minimum or, for that matter, a minimum at all (57a, 57b, 57c). Nevertheless, the simplifying assumption is that minimum energy conformations are at the very least good starting points for examining potential drug candidates. To alleviate the concern of not finding the global minimum, automated conformational search procedures have been devised. In general, one or more bonds may be rotated and evaluated based on some criterion, such as energy differences and/or van der Waals interactions. From these searches, a series of compounds can be saved in a computer file, retrieved later, and then subjected to energy optimization procedures. The problem with conformational searching techniques is that they are computationally demanding; and, if the dihedral angle increment is too large, one could miss some low-energy conformers. The use of conformational searching formed the basis of the active analog approach championed by Marshall (57a, 57b, 57c).

Molecular simulations

Dynamics: Molecular mechanics calculations treat molecules as static, time-averaged structures. Since there is an exponential growth in the number of conformations with increasing degrees of freedom in a molecule, a systematic search of the configuration space for the low-energy conformations of large molecules is virtically impossible. Another powerful computational method which has been applied to conformational problems is molecular dynamics (58a, 58b, 58c). Molecular dynamics (MD) simulations describe the trajectory (configurations as a function of time) of a system by integration of Newton's equations of motion, where the force on particle i is represented as F_i. The phase space trajectories generarated with MD calculations are analyzed to yield equilibrium and dynamical information. See Equations (14a, 14b, 14c, 14d, 14e, 14f, 14g) and (15), where r represents the position and E the potential energy.

$$\frac{d^2 r_i}{dt^2} = m_i^{-1} F_i \tag{14}$$

$$F_i = -\frac{\partial}{\partial r_i} E(r_1, r_2, r_3, \ldots, r_N) \tag{15}$$

This technique, provided a sufficiently good force field has been developed, has certain advantages over molecular mechanics calculations by themselves in that solvation interactions may be determined and different energy minima can be located.

Monte Carlo simulations: The use of Monte Carlo (MC) methods (59) which incorporate the metropolis algorithm has been used the basis for conformational searching (60), drug–receptor docking (61), and free energy perturbation calculations (see next section) (62). MC is based on the principles of statistical mechanics, where an ensemble of molecular states are generated and evaluated. The metropolis MC method for finding available conformations are summarized in the following steps. First, the starting molecular geometry is calculated using a suitable molecular mechanics potential. Second, a small random perturbation is applied, i.e., one of the atoms is kicked from its original position to a new position. Third, a new energy is calculated. If the energy is lower, then the move is accepted, and the new configuration or conformation (which depends on whether the solvent or solute is being perturbed) is stored for later use. If the energy is higher there is a probability that it may or may not be rejected. The probability is based on a Boltzmann factor($e^{-\Delta E/kT}$), where ΔE is the energy difference between the final and original states ($E_2 - E_1$), k is the Boltzmann constant, and T is the temperature. It is necessary to accept some higher-energy-state configurations or conformations to avoid becoming trapped in a local minimum. The procedure described is repeated for many iterations until the lowest free energy is achieved. MC, like MD simulations, can be used to treat solvation.

Free energy perturbations: One of the challenges facing medicinal chemists is to predict (and ultimately prepare and test) compounds with high affinity and efficacy for target receptor sites. Knowing the free energy of binding, $\Delta G^{\circ}_{binding}$ for a series of structurally related compounds is important information. Unfortunately, determining the $\Delta\Delta G_{\circ binding}$ for many different analogs is experimentally demanding. Free energy perturbation

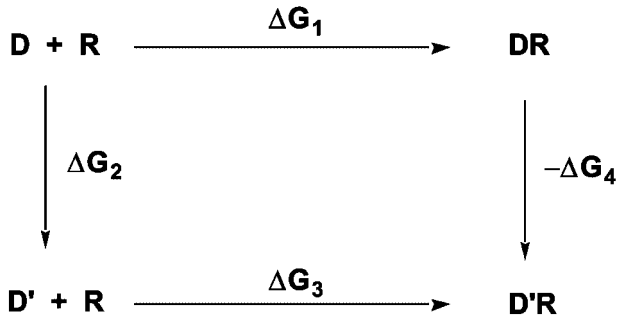


Fig. 2 Free energy perturbation (FEP) diagram, showing drugs D and D′ binding to receptor R to give the DR and DR′ receptor complex and the free energy changes.

(FEP) calculations (63a, 63b) take advantage of the fact that the change in a complete thermodynamic cycle should be zero (see Fig. 1). By measuring the equilibrium constants K_2 and K_1, the experimental $\Delta\Delta G^{\circ}_{binding}$ can be obtained according to Eq. 16. Inspection of Fig. 2, however, reveals that if D_1 could be mutated to D_2 and DR_1 could be mutated into DR_2, we would have a relationship where knowing the difference between $\Delta^{**}G^0_4 - \Delta G^0_3$ would be equivalent to $\Delta G^0_2 - \Delta G^0_1$ (Eq. 17). The former relationship can be simulated using either MD or MC methods.

$$\Delta\Delta G^0_{binding} = \Delta G^0_2 - \Delta G^0_1 = -RT \ln \frac{K_2}{K_1} \tag{16}$$

$$\Delta G^0_2 - \Delta G^0_1 = \Delta G^0_4 - \Delta G^0_3 \tag{17}$$

FEP is a major conceptual achievement with important ramifications for drug design. Unfortunately, two major problems have prevented FEP from being routinely applied to drug design outside of academic research. First, the mutations must be small and gradual. Calculations need to be carried out during the mutations process. Second, the calculations are computationally demanding even on high speed workstations. Nevertheless, it is an attractive computational approach.

COMPUTER-ASSISTED DRUG DESIGN

Computational chemistry applied to drug design relies on two major experimental sources of detailed three-dimensional (3-D) data: 1) X-ray crystallographic analysis and 2) nuclear magnetic resonance (NMR) spectroscopy.

High precision X-ray crystallographic studies now are reported routinely in the literature for small molecules and for proteins and nucleic acids (64). These data provide spatial locations (3-D coordinates) for the atoms in a molecular structure. Since X-rays are diffracted by electron density, hydrogen atoms are not generally located with the same accuracy as heavy atoms. In general, the crystallographer is able to provide complete structures. In protein X-ray crystallography the knowledge of the amino acid sequence of the protein assists the crystallographer in developing a model of the atomic positions and bonding (65). Backbone traces based on three-dimensional electron density make the viewing easier for macromolecular structures. Interestingly, computer-based approaches are also used to fit experimentally determined electron density maps to proposed structures.

NMR techniques are also being used to determine 3-D macromolecular conformation. High-resolution NMR combined with pulse experiments can provide connectivity information along a protein chain. Nuclear overhauser effect (NOE) signals are sensitive to the through-space distances between atoms, and they can be used to determine the folding pattern in a protein. The combination of connectivity data, NOE-determined distance constraints, and amino acid sequence information provides a molecular modeler with a substantial set of geometric constraints for building models of macromolecules. These constraints can be combined to build a macromolecular model using the mathematical technique of distance geometry.

Distance geometry provides sets of 3-D structures of a protein or nucleic acid that fulfill the constraints. The combination of distance geometry, for generation of molecular starting points, with molecular dynamics computations can yield 3-D models of small proteins with precision equal to X-ray crystallography (66). This combination of NMR, molecular mechanics, and molecular dynamics (67) can be used to provide a three-dimensional protein structure in a situation where the protein cannot be crystallized or the crystals are not appropriate for X-ray crystallography.

The term CADD has been used to describe two aspects of the recent use of computational tools that aid computational and medicinal chemists in the search for new drug candidates. In the first approach, medicinal chemists attempt to describe the predominant statistical correlation of biological activity with directly measurable physicochemical parameters or characteristics of drugs and is known as Quantitative Structure-Activity Relationships (QSAR) (68). The central idea is that compounds exhibit biological activity based on structural characteristics. It should then be possible to correlate the associated biological activity with various critical parameters. In general, the biological activity may be considered a function of hydrophobicity, electrostatics, and steric forces (Eq. 18).

$$\text{Biological Activity} = f(\text{hydrophobicity}) + f(\text{electrostatics}) + f(\text{sterics}) \quad (18)$$

The seminal work of Corwin Hansch initiated the field of QSAR (69). Two-dimensional and three-dimensional QSAR methods (2-D QSAR and 3-D QSAR) have been widely applied to problems of biological interest. The latter approach has increased in popularity with the introduction of the comparative molecular field analysis (CoMFA) (70) and commercial availability of similar methods.

The second aspect of CADD relates more closely to computer modeling. This approach can be characterized as pharmacophore mapping, which uses structural information from receptors or small molecules (in combination with computational chemistry techniques) to develop new candidate drug targets. A pharmacophore, first defined by Paul Erlich, is the set of essential molecular features in a compound responsible for a specific pharmacological action. This approach has also been called receptor-based drug design, using a very general definition of the word receptor (71), which broadens the definition of classical pharmacology. In this technique, a receptor or a model of a receptor is used as the target for the drug design efforts. This receptor model can be an X-ray crystallographic structure or a model built from the X-ray structure of a homologous protein. These can be viewed as recognition sites, since it could be argued that the receptor *recognizes* certain patterns of geometry and charge density of a drug as it approaches and binds to its receptor. As improved computational and X-ray structural tools become available, modeling of pharmacophores in 2-D as well as in 3-D, also an old technique, has become a more commonly used and much more sophisticated approach.

Some early pharmacophore models contained sets of hexagonal rings (72) presented as a single plane. These rings were used to describe, in a 2-D fashion, the relative 3-D relationship of receptor functional groups. If no receptor surfaces are known to exist in a specific orientation, this area is called a "bulk tolerance area" of the unknown, but complex, 3-D area of an enzyme's active site (73). Another later example of pharmacophoric models is the triangular N O O hypothesis which Cheng (74a, 74b) used to describe, in simple geometric terms, the conformation of active antitumor agents. This hypothesis used early results from X-ray crystal studies on small molecules and measurements from physical molecular models to suggest the required geometry of active antitumor agents. Through-space distances were used to connect a basic nitrogen with two hydrogen-bond-donating oxygen groups. This pharmacophore model was more sophisticated than most previous models because it was developed with three-dimensional data.

More recently CADD software, for example the Catalyst program from MSI (www.msi.com) (75, 75), is available for pharmacophore modeling and searching of 3-D databases. The use of such a program, only one of many with similar power, shows the increasing sophistication of computational search strategies for drug design and discovery. The input to this program is a structure activity set of molecules. The program then analyzes the set of molecules and generates a set of pharmacophoric hypotheses that can be represented in 3-D. The

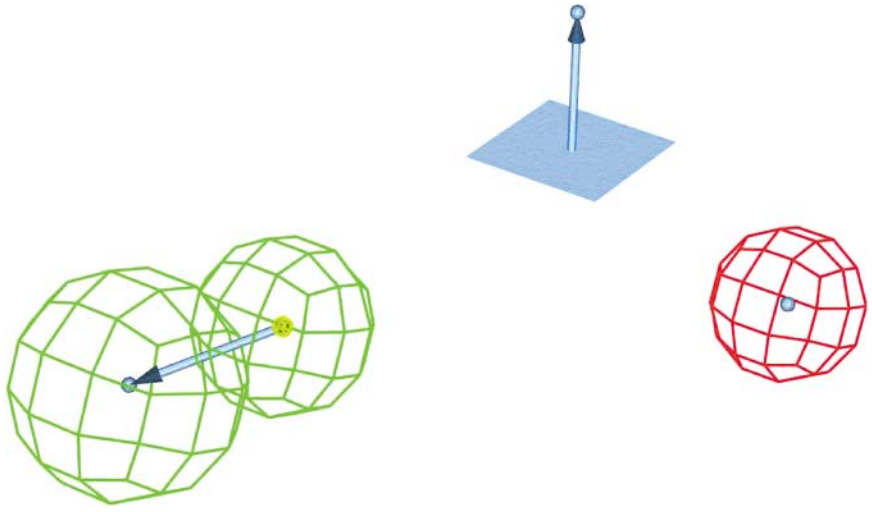


Fig. 3 Three-dimensional pharmacophore generated in catalyst.

pharmacophoric hypotheses can then be used by catalyst to search databases of molecule for pharmacophore matches.

Fig. 3 represents an example of a 3-D pharmacophoric hypothesis as generated in catalyst. This is a 2-D snapshot of the 3-D pharmacophores. They would have inter-pharmacophore distances associated with the computer data object. The blue plane represents the location of a plane of an aromatic ring with the necessary perpendicular vector for orienting aromatic ring systems in new molecules. The smaller of the green spheres represents a hydrogen bond acceptor, sized at 1.7 Å, and the larger sphere representing the projected point and locations for a possible hydrogen atom donating group with a sphere of 2.3 Å. The red sphere represents a positively charged group with a radius of 1.5 Å. This type of pharmacophore hypothesis can then be used as a 3-D search query that, in many cases has proven useful in finding molecules related to the initial input set.

As previously mentioned, X-ray crystallography provides the bulk of the 3-D information used by medicinal and computational chemists in structure based drug design. Two research-based organizations (76, 77) have compiled and provided a subscription service to databases which are the primary source for X-ray structural data. Besides the information contained in the databases, the existence of these organizations and defined computer formats for crystallographic data has the advantage of providing the scientist with some standardization. Commercial molecular modeling software packages provide a transparent interface between graphics display programs and these data. The two databases divide their holdings on the basis of the size of the molecule. Large polynucleotide crystal structures and structures with drugs bound to polynucleotides are stored in both the protein data bank (PDB http://www.rcsb.org/pdb/) and Cambridge files, with the division basically being subjective. Generally, the larger structures are in the PDBdatabase.

The Cambridge Crystallographic Data Center (http://www.ccdc.cam.ac.uk/) (76) now provides a computerized database of more than 200,000 small molecular structures. In addition to this extensive structural database, the Cambridge group has developed a suite of programs for systematic structure search and retrieval. The most unique aspect of these programs is that database search queries can specify desired structures in three dimensions.

With the rapid advances in macromolecule crystallography, the PDB (http://www.rcsb.org/pdb/) (77) currently contains about 12,600 macromolecular NMR or crystal structures. While this may not represent all macromolecular structures, it does represent the large majority. Most of the major scientific journals require simultaneous deposition of macromolecular crystal data into the PDB as the journal article is published. Besides structural proteins, and the classic crystal structures such as hemoglobin and lysozyme, the data base includes many enzymes or representative enzymes from classes that could be good targets for chemotherapy.

Table 1 shows a realistic overview of the successes of CADD in the pharmaceutical industry. Target proteins have been vigorously explored as indicated by the number of structures in the PDB. The only criteria of success that can be used in this situation is to measure compounds put into clinical trials, since about 80% of candidate drugs fail in clinical trials many of these compounds may not end up on the market.

Automated X-ray diffractometers integrated with fast computer workstations are commercially available for rapid acquisition of protein diffraction data. Area detectors for X-ray diffraction allow for rapid collection of much larger quantities of reflection data and provide the ability to find structures for materials that were too unstable to withstand the older, slower techniques. New automated software routines that use direct methods allow companies to provide a commercial service solving crystal structures. Increasing use of synchrotrons around the world has

Table 1 Ouruiew of the successer of CADD

Target (# in PDB)	Research group	Method	Outcome
Thrombin (105)	Hoffman-La Roche	X-ray, models (2.2 Å)	Clinical candidate
Thrombin	Biogen Inc.	X-ray, models	Phase III
Neuraminidase (53)	Monash University/GW	GRID search, X-ray (2.4 Å)	Phase II*
Purine nucleoside phosphorylase (16)	Biocryst	Molecular mechanics X-ray (1.5–2.0 Å)	Phase II
Thymidylate synthase (66)	Agouron	Modeling X-ray (1.9–2.5 Å)	Clinical candidate
Carbonic Anhydrase (111)	Merck	Multiple X-ray (1.9–2.5 Å)	Market
Rhinovirus-14 (28)	Sterling Winthrop	Screening X-ray (2.9 Å)	Phase I
Aldose reductase (13)	Ayerst	Molecular orbital QSAR (1.8–2.3 Å)	Market

allowed rapid collection of X-ray data. The technology has expanded widely into the pharmaceutical industry. Consortia such as IMCA provide drug discovery with a significant resource for crystallography. (http://www.imca.aps.anl.gov/)

STATE-OF-THE-ART COMPUTER-ASSISTED DRUG DESIGN

Theory

Most recently an interesting model for approaching CADD has been presented (78), where a set of guidelines for the practical application of CADD techniques are provided. It relates the number of compounds that need to be investigated in a drug-discovery project to the amount of information available about the drug target. More information, available about the target, means that fewer compounds must be investigated in the drug design process and also changes the techniques used by the computational chemist. This is illustrated graphically in Fig. 4.

On the one hand, if the protein X-ray crystallographic structure has been characterized, then the techniques of CADD are used directly. This would involve use of the crystal structure as a target of design efforts using the structural information. That initial effort would be followed by solving additional structures with new, more potent, ligands bound to the protein. The bound structures are modified, graphically, and then evaluated using methods previously discussed. Only compounds that show promise as possible ligands are actually prepared. On the other hand, if there is less information about the target available: 1) a protein sequence and no characterized structure or 2) a set of compounds with different degrees of biological activity. Then the pharmacophoric model and hypothesis approach, discussed above, can be used. A 3-D pharmacophoric model building effort, using

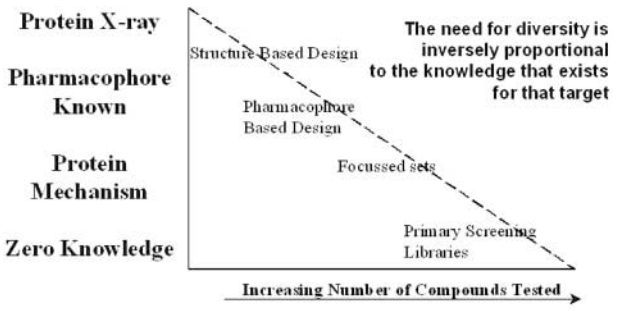


Fig. 4 Plot of relationship between information and number of compounds necessary for synthesis.

catalyst might develop pharmacophore models that can be used to search libraries of structures.

At the other extreme if you have no knowledge of the protein structure, no candidate ligands, and the X-ray structure is not yet known then a high throughput screening of primary large scale collections of compounds or combinatorial libraries would be screened. These techniques would use roboticized and miniaturized assays and be done to search for possible binding compounds.

Methodology

Design of enzyme inhibitors guided by molecular modeling and X-ray crystallography

An excellent, and successful, structure based drug design example in a system where there is significant structural information is work on the development of new inhibitors of HIV protease (79a, 79b). This work was carried out by scientists from Dupont Merck Pharmaceuticals. A substantial amount of structural data has been available for a decade on the structure of HIV protease (80).

HIV protease is one of the proteins coded by the HIV viral genome and expressed as part of the reproductive cycle of the virus. HIV protease contains active aspartic acid residues that classify it as an aspartyl protease. The viral genome once incorporated into the host genome codes for a polyprotein. The protease is responsible for cleaving specific sites in the polyprotein into specific proteins that allow the virus to mature. The HIV protease is a symmetrical dimer of with each monomer containing 99 amino acids. Fig. 5 shows a cartoon of the protease structure with the active site region labeled. This region is responsible for the binding of the protein substrate sequence. It recognizes and binds a hexapeptide and hydrolyzes the labile central peptide bond.

Fig. 6 shows the binding mode of an HIV protease inhibitor from a X-ray study (file 4phv). In this figure two amino acids, glycine 49 (GLY49) and isoleucine 50 (ILE50) of the HIV protease, are shown in stick form. The remainder of the protein-active site is in line form. Since the enzyme is a symmetrical dimer, the residues from each dimer are shown although only one pair of residues is labeled. Also shown is a water molecule bound in the active site of the enzyme. It is believed that this water molecule assists in the peptide bond hydrolysis. The protein NH groups bind to this water molecule. Carbonyl groups of the inhibitor also interact with this water molecule. Additionally this crystal structure shows the position of the peptidomimetic, HIV protease inhibitor, **1**, and the manner in which the four aromatic ring systems fit into the enzyme.

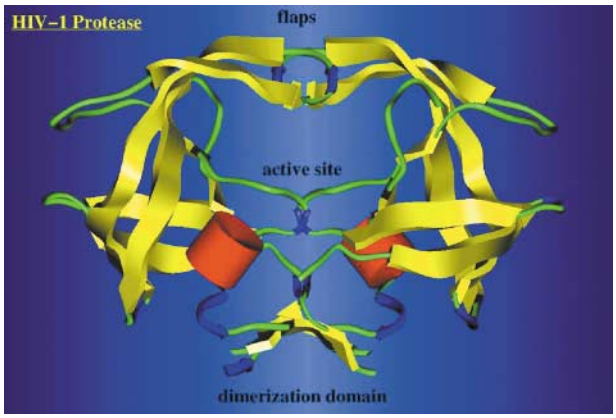


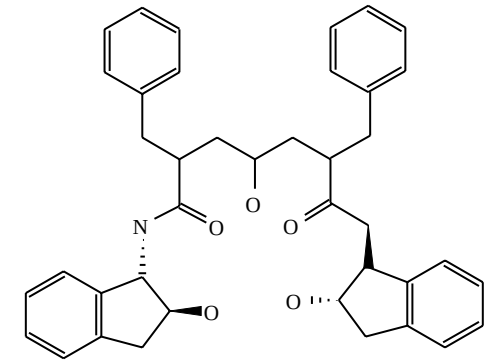
Fig. 5 HIV protease structure with active site region labeled.

The second structure in Fig. 6 represents a similar view of the crystal structure (file lajx). In this structure-based design, a cyclic urea 2 was synthesized in which it was hypothesized that the water molecule could be replaced by a functional group of the inhibitor (81). In the case of compound 2, the oxygen of the water was replaced by the oxygen of a carbonyl group that interacts with the backbone NH groups just as the water molecule does. The four aromatic groups of the cyclic urea structure.

Equipment

Currently, molecular modeling hardware is concentrated on two groups of computer systems. High-end workstations with high-resolution graphics display systems for molecular modeling are primarily Silicon Graphics models, running the Silicon Graphics version of the UNIX operating system. These systems have held predominance in the market for about a decade. A rapidly growing competitor is the standard Intel chip powered, PC running Windows or Windows/NT. The vast majority of the pharmaceutical industry has supported Windows/NT desktops for scientists. Marketing efforts by computational chemistry software vendors have been shifting rapidly to that arena.

While there is considerable press about the Linux version of UNIX, it has still not penetrated the computational chemistry market to a large extent. Future high end computing will probably depend upon versions of UNIX with transparent delivery to desktop workstations.



Structure 1.

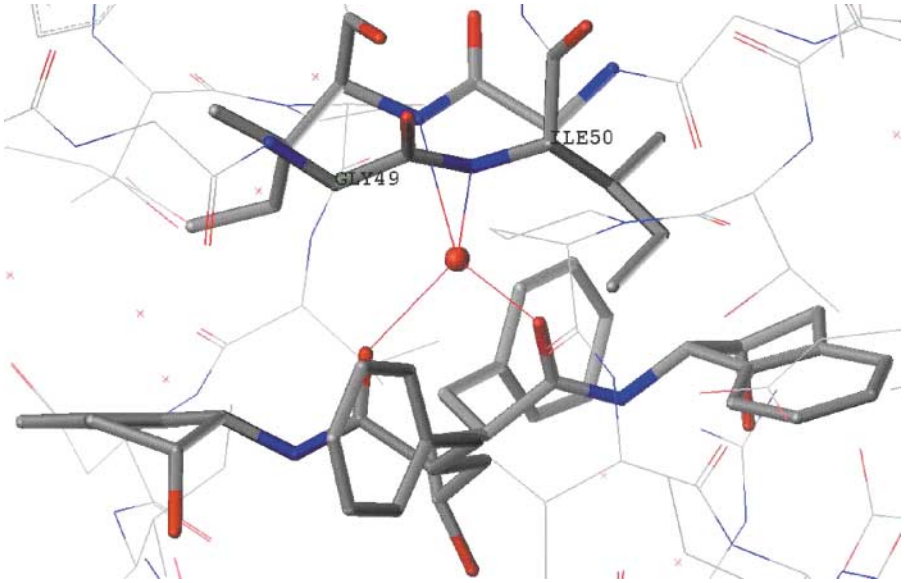


Fig. 6 X-ray structure of HIV protease and protease inhibitor taken from PDB.

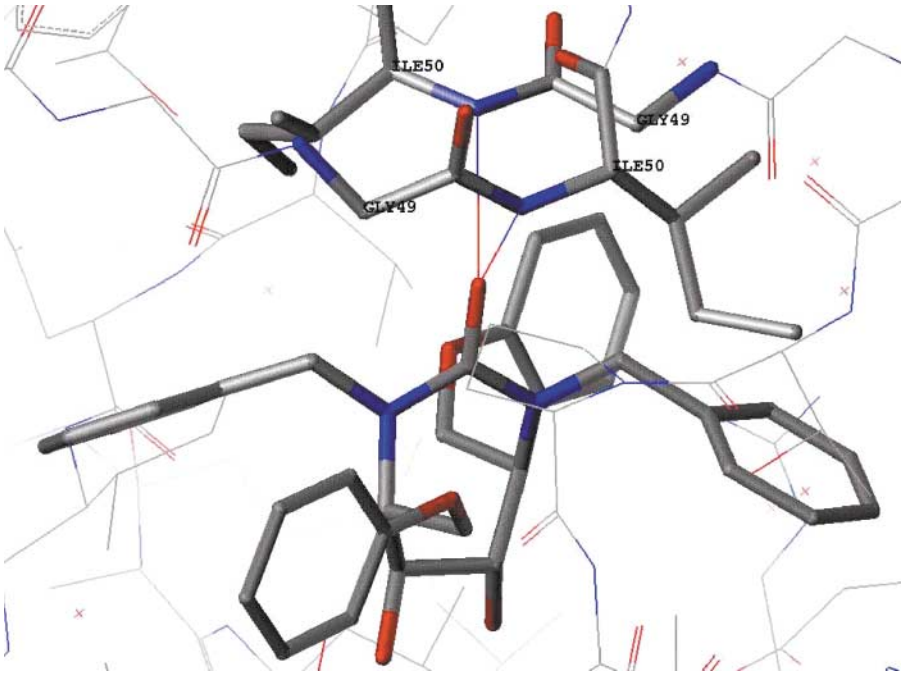
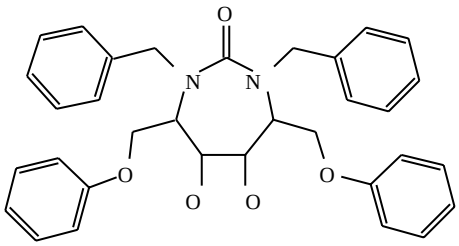


Fig. 7 X-ray structure of cyclic HIV protease inhibitor with carbonyl group mimic of water.

Software

The emergence of molecular modeling and computational into a discipline has been closely associated with advances in computer hardware, computer graphics, and software. The availability of more sophisticated number-crunching power has allowed scientists to calculate the properties of larger and larger molecules. With the advent of supercomputers, dynamics calculations on biomolecules surrounded by an environment of water are becoming common.

The advances in software have also been spectacular. It was not terribly long ago when medicinal scientists were happy to get a color picture that somehow conveyed a three-dimensional perspective. Now software packages allow chemists to conveniently overlay a series of structures and do real-time rotations. Many companies have emerged, most originated by academicians, which specialize in designing and marketing modeling software. The origins of much of the available software can be traced to university research laboratories. Although many of the packages have similar features and capabilities, they each have a characteristic style, which reflects the philosophies of the developers.



Structure 2.

CONCLUSIONS AND FUTURE TRENDS

Chemists have long been associated with ball-and-stick mechanical structures, and the concepts for molecular modeling have been appreciated for a number of years by medicinal chemists. The precomputer modeling endeavors of Pauling and Watson and Crick demonstrated the utility of these methodologies. The mathematical formalisms for quantum and molecular mechanics were worked out years ago; however, the lack of sufficient computational means prevented practical applications. The explosion of modeling endeavors today is directly related to the advent of powerful new computers and algorithms. Now, larger and larger structures can be subjected readily to the rigors of calculations. More realistic simulations, including solvent interactions, can be treated. Molecular structures can be easily displayed and manipulated on computer graphics terminals. The full impact of high-speed desktop workstations, the Internet and the genome initiative.

Computational chemists, medicinal chemists, and pharmaceutical scientists face an exciting future, as the computer-assisted approaches for drug design are becoming more powerful and more accessible. Every major pharmaceutical company has invested resources into these new tools. Education in the techniques, scope, and limitations are ever increasing. Improvements in the methodologies and computational resources will undoubtedly increase. Nevertheless, biological processes are still poorly understood, for the most part, at the molecular level. Many challenges remain, and theoretical and experimental work will remain tightly correlated in drug design and discovery.

REFERENCES

1. Wooden Ball-and-Stick Models Attributed to John Dalton. The Science Museum, Exhibition Road, South Kensington: London, England.
2. Watson, J.D. *The Double Helix*; New American Library: New York, 1968.
3. Watson, J.D.; Crick, F.H.C. Nature **1953**, *171*, 964–967.
4. Gordon, A.J. J. Chem. Educ. **1970**, *47*, 30–33.
5. Koltun, W. Biopolymers **1965**, *3*, 665–679.
6. Harvard Apparatus Inc.: South Natick, Mass, 01760.
7. Johnson, C.K. Report No. ORNL-3794, 1970 Oak Ridge National Laboratory: Oak Ridge, Tennessee.
8. Connolly, M.L. Science **1983**, *221*, 709–713.
9. Fuchs, H.; Goldfeather, J.; Hultquist, J.P.; Spach, S.; Austin, J.D.; Brooks, F.P., Jr.; Eyles, J.G.; Poulton, J. Computer Graphics **1985**, *19*, 111–120.

10a. D'Abro, A. *The Rise of the New Physics: Its Mathematical and Physical Theories*; Dover: New York, 1951.

10b. Pauling, L.; Wilson, E.B., Jr. *Introduction to Quantum Mechanics, With Applications to Chemistry*; McGraw-Hill: New York, 1935.

11a. Schrödinger, E. Ann. Phys. **1926**, *79*, 361.

11b. Schrödinger, E. Ann. Phys. **1926**, *80*, 437.

11c. Schrödinger, E. Ann. Phys. **1926**, *81*, 109.

12. Heisenberg, W. Z. Phys. **1925**, *33*, 879–893.

13a. Schrödinger, E. Ann. Phys. **1926**, *79*, 734.

13b. Eckart, C. Phys. Rev. **1926**, *28*, 711.

14a. Pople, J.A.; Beveridge, D. *Approximate Molecular Orbital Theory*; McGraw-Hill: New York, 1970.

14b. Dewar, M.J.S. *The Molecular Orbital Theory of Organic Chemistry*; McGraw-Hill: New York, 1969.

14c. Murrell, J.N.; Harget, A.J. *Semiempirical Self-Consistent-Field Molecular Orbital Theory of Molecules*; Wiley-Interscience: London, 1972.

14d. Pople, J.A.; Santry, D.P.; Segal, G.A. J. Chem. Phys. **1965**, *43*, S129–S135.

14e. Pople, J.A.; Segal, G.A. J. Chem. Phys. **1965**, *43*, S136–S149.

14f. Pople, J.A.; Segal, G.A. J. Chem. Phys. **1966**, *44*, 3289–3296.

14g. Pople, J.A.; Beveridge, D.L.; Dobosh, P.A. J. Chem. Phys. **1967**, *47*, 2026–2033.

15. Dewar, M.J. J. Mol. Struct. **1983**, *100*, 41–50.
16. Dewar, M.J.S.; Thiel, W. J. Am. Chem. Soc. **1977**, *99*, 4899–4907.

17a. Bingham, R.C.; Dewar, M.J.S.; Lo, D.H. J. Am. Chem. Soc. **1975**, *97*, 1285–1293.

17b. Bingham, R.C.; Dewar, M.J.S.; Lo, D.H. J. Am. Chem. Soc. **1975**, *97*, 1294–1301.

17c. Bingham, R.C.; Dewar, M.J.S.; Lo, D.H. J. Am. Chem. Soc. **1975**, *97*, 1302–1306.

17d. Bingham, R.C.; Dewar, M.J.S.; Lo, D.H. J. Am. Chem. Soc. **1975**, *97*, 1307–1310.

17e. Dewar, M.J.S.; Lo, D.H.; Ramsden, C.A. J. Am. Chem. Soc. **1975**, *97*, 1311–1318.

18a. Dewar, M.J.S.; Thiel, W.J. Am. Chem. Soc. **1977**, *99*, 4907–4917.

18b. Dewar, M.J.S.; McKee, M.L. J. Am. Chem. Soc. **1977**, *99*, 5231–5241.

18c. Dewar, M.J.S.; Rzepa, H.S. J. Am. Chem. Soc. **1978**, *100*, 777–794.

18d. Davis, L.P.; Guidry, R.M.; Williams, J.R.; Dewar, M.J.S.; Rzepa, H.S. J. Comput. Chem. **1981**, *2*, 433–445.

18e. Dewar, M.J.S.; McKee, M.L.; Rzepa, H.S. J. Am. Chem. Soc. **1978**, *100*, 3607.

18f. Dewar, M.J.S.; Healy, E. J. Comput. Chem. **1983**, *4*, 542–551.

19. Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.F.; Stewart, J.J.P. J. Am. Chem. Soc. **1985**, *107*, 3902–3909.

20a. Richards, W.G. *Quantum Pharmacology*; Butterworths: London, 1984.

21. Kier, L.B. *Molecular Orbital Theory in Drug Research*; Academic Press: New York, 1984.
22. Martin, Y.C. *Quantitative Drug Design*; Marcel Dekker, Inc.: New York, 1978.
23. Hehre, W.J.; Random, L.; Schleyer, P.; Pople, J. Ab Initio *Molecular Orbital Theory*; John Wiley & Sons: New York, 1986.
24. Pople, J.A. J. Am. Chem. Soc. **1975**, *97*, 5306–5308.
25. Hehre, W.J. J. Am. Chem. Soc. **1975**, *97*, 5308–5310.
26. Dewar, M.J.S. J. Am. Chem. Soc. **1975**, *97*, 6951.
27. Roberts, J.D. *Molecular Orbital Calculations*; Benjamin/-Cummings: Reading, Massachusetts, 1961.
28. Streitwieser, A., Jr. *Molecular Orbital Theory for Organic Chemists*; John Wiley & Sons: New York, 1961.
29. Boys, S.F. Proc. R. Soc. (London A) **1950**, *200*, 542–554.
30. Hehre, W.J.; Stewart, R.F.; Pople, J.A. J. Chem. Phys. **1969**, *51*, 2657–2664.
31. Hehre, W.J.; Ditchfield, R.; Stewart, R.F.; Pople, J.A. J. Chem. Phys. **1970**, *52*, 2769–2773.
32. Pietro, W.J.; Levi, B.A.; Hehre, W.J.; Stewart, R.F. Inorg. Chem. **1980**, *19*, 2225–2229.
33. Pietro, W.J.; Blurock, E.S.; Hout, R.F., Jr.; Hehre, W.J.; DeFrees, D.J.; Stewart, R.F. Inorg. Chem. **1981**, *20*, 3650–3654.
34. Binkley, J.S.; Pople, J.A.; Hehre, W.J. J. Am. Chem. Soc. **1980**, *102*, 939–947.
35. Gordon, M.S.; Binkley, J.S.; Pople, J.A.; Pietro, W.J.; Hehre, W.J. J. Am. Chem. Soc. **1982**, *104*, 2797–2803.
36. Dobbs, K.D.; Hehre, W.J. J. Comput. Chem. **1986**, *7*, 359–378.
37. Hehre, W.J.; Ditchfield, R.; Pople, J.A. J. Chem. Phys. **1972**, *56*, 2257–2261.

38. Binkley, J.S.; Pople, J.A. J. Chem. Phys. **1977**, *66*, 879–880.
39. Francl, M.M.; Pietro, W.J.; Hehre, W.J.; Binkley, J.S.; Gordon, M.S.; DeFrees, D.J.; Pople, J.A. J. Chem. Phys. **1982**, *77*, 3654–3665.
40. Carlsen, N.R. Chem. Phys. Lett. **1977**, *51*, 192–196.
41. Krishnan, R.; Frisch, M.J.; Pople, J.A. J. Chem. Phys. **1980**, *72*, 4244–4245.
42. Hartree, D.R. Proc. Cambridge Philos. Soc. **1928**, *24*, 105.
43. Roothaan, C.C.J. Rev. Mod. Phys. **1951**, *23*, 69.
44. Hall, G.G. Proc. roy. soc. (london a) **1951**, *205*, 541.
45. Møller, C.; Plesset, M.S. Phys. Rev. **1934**, *46*, 618–622.
46. Pople, J.A.; Binkley, J.S.; Seeger, R. Int. J. Quantum Chem. Simp. **1976**, *10*, 1–19.
47. Pople, J.A.; Seeger, R.; Krishnan, R. Int. J. Quantum Chem. Quantum Chem. Simp. **1977**, *11*, 149–163.
48. Krishnan, R.; Pople, J.A. Int. J. Quantum Chem. **1978**, *14*, 91–100.
49. Schlegel, H.B. J. Chem. Phys. **1982**, *77*, 3676–3681.
50. Parr, R.G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, 1989.
51. Kohn, W.; Sham, L.J. Phys. Rev. **1934**, A*140*, 1133–1138.
52. Hohenberg, P.; Kohn, W. Phys. Rev. **1964**, B*136*, 864–871.
53. Burkert, U.; Allinger, N.L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.
54. Westheimer, F.H.; Mayer, J.E. J. Chem. Phys. **1946**, *14*, 733–738.
55. Hill, T.L. J. Chem. Phys. **1946**, *14*, 465.
56. Dostrovsky, I.; Hughes, E.D.; Ingold, C.K. J. Chem. Soc. **1946**, *173–194.*
57. Boyd, D.B. Drug Inf. J. **1983**, *17*, 121–131.
58. Goldstein, H. *Classical Mechanics*; Addison-Wesley: Reading, Massachusetts, 1981.
59. Allinger, N.L. J. Am. Chem. Soc. **1977**, *99*, 8127, and subsequent papers; MM2.
60. Allinger, N.L.; Yuh, Y.H.; Lii, J.-H. J. Am. Chem. Soc. **1989**, *111*, 8551, MM3.
61. Lii, J.-H.; Allinger, N.L. J. Am. Chem. Soc. **1989**, *111*, 8566.
62. Lii, J.-H.; Allinger, N.L. J. Am. Chem. Soc. **1989**, *111*, 8576.
63. Bowen, J.P.; Allinger, N.L. Molecular Mechanics: The Art and Science of Parameterization. *Reviews Computational Chemistry*; Lipkowitz, K., Boyd, D., Eds.; VCH: New York, 1991; 2, 81–97.
64. Allinger, N.L.; Chen, K.S.; Lii, J.-H. J. Comp. Chem. **1996**, *17*, 642, MM4.
65. Nevins, N.; Chen, K.S.; Allinger, N.L. J. Comp. Chem. **1996**, *17*, 669.
66. Nevins, N.; Lii, J.-H.; Allinger, N.L. J. Comp. Chem. **1996**, *17*, 695.
67. Nevins, N.; Allinger, N.L. J. Comp. Chem. **1996**, *17*, 730.
68. Allinger, N.L.; Chen, K.S.; Katzenellenbogen, J.A.; Wilson, S.R.; Anstead, G.M. J. Comp. Chem. **1996**, *17*, 747.
69. Bowen, J.P. New Vistas in Molecular Mechanics Calculations. *Drug Design*; Charifson, P., Ed.; Marcel Dekker, Inc.: New York, 1997; 495–538.
70. Lennard-Jones, J.E. Proc. Roy. Soc. (London, A). **1924**, *106*, 463.
71. Bischoff, C.A. Ber. Dtsch. Chem. Ges. **1890**, *23*, 620, 1891, 24, 1074, 1086; 1891, 26, 1452.
72. Christie, G.H.; Kenner, J. J. Chem. Soc. **1922**, *121*, 614–620.
73. Kemp, J.D.; Pitzer, K.S. J. Chem. Phys. **1936**, *4*, 749.
74. Kemp, J.D.; Pitzer, K.S. J. Am. Chem. Soc. **1937**, *59*, 276–279.
75. Pitzer, R.M. Acc. Chem. Res. **1983**, *16*, 207–210.
76. Bartell, L.S. J. Am. Chem. Soc. **1977**, *99*, 3279–3282.
77. Allinger, N.L.; Hindman, D.; Honig, H. J. Am. Chem. Soc. **1977**, *99*, 3282–3284.
78. Radom, L.; Hehre, W.J.; Pople, J.A. J. Am. Chem. Soc. **1972**, *94*, 2371–2381.
79. Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P.J. J. Am. Chem. Soc. **1984**, *106*, 765–784.
80. Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. J. Comp. Chem. **1986**, *7*, 230–252.
81. Pearlman, A.; Case, D.A.; Caldwell, J.W.; Ross, W.S.; Cheatham, T.E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Comput. Phys. Commun. **1995**, *91*, 1.
82. Cornell, D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. J. Am. Chem. Soc. **1995**, *117*, 5179.
83. Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. J. Am. Chem. Soc. **1984**, *106*, 765–784.
84. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; Stales, D.J.; Swaminathan, S.; Karplus, M. J. Comput. Chem. **1983**, *4*, 187–217.
85. Buyong, M.; Lii, J.-H.; Allinger, N.L. J. Comp. Chem. **2000**, *21*, 814–825.
86. Bowen, J.P.; Pathiaseril, A.; Profeta, S., Jr.; Allinger, N.L. J. Org. Chem. **1987**, *52*, 5162–5166.
87. Bowen, J.P.; Allinger, N.L. J. Org. Chem. **1986**, *51*, 1513–1516.
88. Goldsmith, D.J.; Bowen, J.P.; Quamhiyeh, E.; Still, W.C. J. Org. Chem. **1987**, *52*, 951–953.
89. Bowen, J.P.; Allinger, N.L. J. Org. Chem. **1987**, *52*, 1830–1834.
90. Bowen, J.P.; Allinger, N.L. J. Org. Chem. **1987**, *52*, 2937–2938.
91. Bowen, J.P.; Reddy, V.; Patterson, D., Jr.; Allinger, N.L. J. Org. Chem. **1988**, *53*, 5471–5475.
92. Green, M.M.; Boyle, B.A.; Vairamani, M.; Mukhopadhyay, T.; Saunders, W.H., Jr.; Bowen, P.; Allinger, N.L. J. Am. Chem. Soc. **1986**, *108*, 2381–2387.
93. Spellmeyer, D.C.; Houk, K.N. J. Org. Chem. **1987**, *52*, 959–974.
94. Dorigo, A.E.; Houk, K.N. J. Am. Chem. Soc. **1987**, *109*, 3698–3708, (corrections), 1988, 110, 4874.
95. Rudolf, K.; Hawkins, J.M.; Loncharich, R.J.; Houk, K.N. J. Org. Chem. **1988**, *53*, 3879–3882.
96. Wu, Y.-D.; Houk, K.N. J. Am. Chem. Soc. **1987**, *109*, 908.
97. Jacoby, S.L.S.; Kowalik, J.S.; Pizzo, J.T. *Iterative Methods for Nonlinear Optimization Problems*; Prentice Hall, Englewood Cliffs: New Jersey, 1972.
98. Williams, J.E.; Stang, P.J.; Schleyer, P.V.R. Annu. Rev. Phys. Chem. **1968**, *19*, 531.
99. Wibeg, K.B. J. Am. Chem. Soc. **1965**, *87*, 1070–1078.
100. Fletcher, R.; Reeves, C.M. Comput. J. **1964**, *7*, 149.

101. Fletcher, R.; Powell, M.J.D. Comput. J. **1963**, *6*, 163.
102. Jacoby, S.L.S.; Kowalik, J.S.; Pizzo, J.T. *Iterative Methods for Nonlinear Optimization Problems*; Prentice Hall, Englewood Cliffs: New Jersey, 1972.
103. Ermer, O. Struct. Bonding (Berlin) **1976**, *27*, 161.
104. Marshall, G.R. Medicinal Chemistry VI, Proceeding of the 6th International Symposium on Medical Chemistry. Sinkin, A., Ed.; Cotswold Press: Oxford, 1979; 225–235.
105. Marshall, G.R.; Motoc, I. Top. Mol. Pharmacol. **1986**, *3*, 115–155.
106. Marshall, G.R. Annu. Rev. Pharmacol. Toxicol. **1987**, *27*, 193–213.
107. Beveridge, D.L., Computer Simulation of Chemical and Biochemical Systems; Jorgensen, W.L., Eds.; New York, 1986.
108. McCammon, J.A.; Harvey, S.C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University: Cambridge, 1987.
109. van Gunsteren, W.F., Weiner, P.K., Eds.; *Computer Simulation of Biomolecular Systems*; ESCOM: Leiden, 1989.
110. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E.J. Chem. Phys. **1953**, *21*, 1087.
111. Saunders, M.; Houk, K.N.; Wu, Y.-D.; Still, W.C.; Lipton, M.; Chang, G.; Guida, C. J. Am. Chem. Soc. **1990**, *113*, 1419–1427.
112. Miranker, A.; Karplus, M. J. Am. Chem. Soc. **1990**, *112*, 9161–9175.
113. Jorgensen, W.L.; Ravimohan, C.J. Chem. Phys. **1985**, *83*, 3050–3054.
114. McCammon, J.A. Science **1987**, *238*, 486–491.
115. Jorgensen, W.L. Acc. Chem. Res. **1989**, *22*, 184–189.
116. Horn, A.S., de Ranter, C.J., Eds. *X-ray Crystallography and Drug Action*; Clearendon: Oxford, 1984.
117. Stezowski, J.J.; Chandrasekhar, K. Annu. Rep. Med. Chem. **1986**, *21*, 293–302.
118. Jefson, M.R. Annu. Rep. Med. Chem. **1988**, *23*, 275–283.
119. Clore, G.M.; Gronenborn, A.M.; Brünger, A.T.; Karplus, M. J. Mol. Biol. **1985**, *186*, 435–455.
120. Martin, Y.C. *Quantitative Drug Design—A Critical Introduction*; Marcel Dekker, Inc.: New York, 1978.
121. Hansch, C. Acc. Chem. Res. **1969**, *2*, 232–239.
122. Cramer, R.D., III; Patterson, D.E.; Bunce, J.D. J. Am. Chem. Soc. **1988**, *110*, 5959–5967.
123. Gund, P. Annu. Rep. Med. Chem. **1978**, *14*, 299–308.
124. Baker, B.R.; Wood, W.F. J. Med Chem. **1968**, *11*, 644–649.
125. Baker, B.R. *Design of Active-Site-Directed Irreversible Enzyme Inhibitors*; Wiley: New York, 1967.
126. Zee-Cheng, K.-Y.; Cheng, C.C. J. Pharm. Sci. **1970**, *59*, 1630–1634.
127. Townsend, L.B.; Schram, K.H.; Beranek, J. Proc. Natl. Acad. Sci. USA. **1979**, *76*, 4075–4078.
128. Smellie, A.; Kahn, S.D.; Teig, S.L. J. Chem. Inf. Comput. Sci. **1995**, *35* (2), 295–304.
129. Smellie, A.; Kahn, S.D.; Teig, S.J. Chem. Inf. Comput. Sci. **1995**, *35* (2), 285–294.
130. Allen, F.H.; Bellard, S.; Brice, M.D.; Cartwright, B.A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B.G.; Kennard, O.; Motherwell, W.D.S.; Rogers, J.R.; Watson, D.G. Acta Crystallogr. Sect. B **1979**, *35*, 2331–2339.
131. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.B.; Meyer, E.F., Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. J. Mol. Biol. **1977**, *112*, 535–542.
132. Hann, M.; Green, R. Chemoinformatics Chem. Biol. **1999**, *3* (4), 379–383.
133. Swanstrom, R.; Erona, J. J. Pharmacol. Ther. **2000**, *86* (2), 145–170.
134. Leung, D.; Abbenante, G.; Fairlie, D.P. J. Med. Chem. **2000**, *43* (3), 305–341.
135. Wlodawer, A.; Erickson, J.W. Annu. Rev. Biochem. **1993**, *62*, 543–585.
136. Lam, P.Y.S.; Jadhav, P.K.; Eyermann, C.J.; Hodge, C.N.; Ru, Y.; Bacheler, L.T.; Meek, J.L.; Otto, M.J.; Rayner, M.M.; Wong, Y.N.; Chang, C.-H.; Weber, P.C.; Jackson, D.A.; Sharpe, T.R.; Erickson-Viitanen, S. Science **1994**, *263*, 380–384.